Development of measurement instruments for 21st century skills of elementary school students based on the STEAM approach: A study of validity and reliability

Melva Zainil[1], Ary Kiswanto Kenedi[2*], Yesi Anita[3], Asna Mardin[4], Nova Yulistina[5]

[1]Primary School Teacher Education Study Program, Universitas Negeri Padang, Padang, 25131, Indonesia.
[2]Primary School Teacher Education Study Program, Universitas Samudra, Langsa, Indonesia.
[3]Doctoral Program in Educational Science, Universitas Negeri Padang, Padang, Indonesia.
[4]SDN 19 Paninjauan, Department of Education and Culture, Tanah Datar Regency, Indonesia.
[5]Master's Program in Educational Technology, Universitas Negeri Padang, Padang, Indonesia.

Corresponding author: Ary Kiswanto Kenedi (*Email: arykenedi@unsam.ac.id*)

Abstract

This study was conducted to develop a valid and reliable instrument to measure the 21st-century skills of elementary school students in the context of STEAM-based learning. The 21st-century skills include five main dimensions: critical thinking, creativity, collaboration, communication, and digital literacy. The research used a Research and Development (R&D) approach with stages including construct development, expert content validation, field testing, construct validity analysis using Confirmatory Factor Analysis (CFA), and reliability testing. The pilot sample consisted of 120 fifth-grade students from six elementary schools in West Sumatra Province. Content validation results showed a Content Validity Index (CVI) value of 0.89, while CFA analysis showed a good measurement model with a CFI value of 0.94 and an RMSEA value of 0.063. All dimensions showed high internal reliability, with Cronbach's Alpha values between 0.76 and 0.83 and Composite Reliability (CR) between 0.81 and 0.87. These results indicate that the developed instrument is suitable for use as a valid and consistent measure of elementary school students' 21st-century skills and can support contextual learning assessments relevant to future education demands.

Keywords: 21st century skills, Construct validity, Elementary school, Measurement instrument, Reliability, STEAM.

1. Introduction

The 21st century is characterized by rapid developments in information technology, the complexity of global issues, and changes in socio-economic dynamics that demand new competencies from every individual [1-5]. In this context, the education system faces the challenge of teaching knowledge and developing critical thinking, creativity, collaboration, communication, and digital literacy skills, known as 21st-century skills [5-12]. As the main foundation for shaping children's character and cognitive abilities, basic education plays a strategic role in preparing students to face a dynamic and uncertain future [13-17].

Several international organizations, such as the Partnership for 21st Century Skills (P21), the World Economic Forum, and UNESCO, emphasize that 21st-century skills must be integrated from the primary education level. P21, for example, identifies four core competencies—critical thinking, communication, collaboration, and creativity (4C)—as essential skills that need to be developed through the education system [18, 19]. In Indonesia itself, the strengthening of character and 21st-century skills has been reflected in the Merdeka Curriculum policy, which promotes a contextual, project-based learning approach and encourages the strengthening of the Pancasila student profile [20-25].

However, the implementation of strengthening 21st-century skills will not be effective without accurate and context-appropriate measurement instruments. A good instrument must be able to measure complex competencies, not only from a cognitive aspect, but also from an affective and psychomotor aspect. Unfortunately, various studies show that the measurement of 21st-century skills, especially at the elementary school level, still faces several limitations. First, many of the instruments used originate from abroad and have not undergone a process of adaptation and validation in the context of Indonesian education [26]. Second, the measurement approach is often purely cognitive, without considering the process and real application in learning [27, 28].

The STEAM (Science, Technology, Engineering, Arts, and Mathematics) approach is a potential alternative to bridge these needs. STEAM is an integrated learning model that combines various disciplines with a project-based approach and contextual problem solving. Through STEAM, students are encouraged to understand concepts and apply them creatively and collaboratively in real-life situations [19, 29-32]. This approach aligns with the development of 21st-century skills, as it encourages continuous exploration, discovery, and communication in the learning process [33-35].

Several studies have stated that the application of STEAM in primary education can improve critical thinking and problem-solving skills and strengthen students' curiosity and collaborative attitudes [10, 36, 37]. However, even though the STEAM approach has been widely used in learning, developing measurement instruments for 21st-century skills based on the STEAM context is still very limited, especially in Indonesia. This indicates a gap between the pedagogical approaches used and the available evaluation tools.

The instruments currently used to measure the 21st-century skills of elementary students are generally still generic, not integrated with the learning approaches applied in the classroom, and do not meet the criteria of scientific validity and reliability . On the other hand, existing studies still focus on middle and high school students, assuming that 21st-century skills do not need to be developed in depth at the elementary level [17, 38-41]. In fact, based on Piaget's cognitive development theory, elementary school age is a transition period from the concrete operational stage to the formal operational stage, which is the ideal moment to instill critical and creative thinking patterns [42-47].

This research gap opens up space for developing new instruments that are not only contextually relevant but can also be validated through modern statistical quantitative approaches, such as confirmatory factor analysis (CFA) or the Rasch model. These instruments need to be carefully designed, starting with competency mapping, indicator development, and item writing and ending with empirical testing of their validity and reliability.

This study aims to develop a 21st-century skills measurement instrument for elementary school students designed based on the STEAM learning approach, which has been empirically tested to ensure its construct validity and reliability. The novelty of this research lies in several aspects. First, this research specifically targets elementary school students as the

main subjects, who have been neglected in developing 21st-century skills instruments. Second, the instrument integrates the STEAM approach directly into its constructs and indicators, so the measurement results better reflect the actual learning process in the classroom. Third, the instrument was comprehensively tested through expert validation, field trials, and quantitative statistical analysis to ensure the methodological quality of the instrument.

Thus, the results of this study are expected to provide practical contributions in the form of measuring tools that can be used by teachers and researchers and theoretical contributions in the field of educational measurement. In addition, this study provides a starting point for the development of contextual and data-driven policies and strategies to strengthen 21st-century skills from the elementary education level.

Through the STEAM approach, basic education can become a space that transmits knowledge and builds the thinking skills needed to face future challenges. However, to ensure that this process runs effectively, appropriate, valid, and reliable measurement instruments are needed. Therefore, developing STEAM-based instruments to measure 21st-century skills is a strategic step in building an adaptive and transformative education evaluation system.

2. Materials and Methods

2.1. Research Design

This research and development (R&D) study aims to produce a measurement instrument for 21st-century skills based on the STEAM approach and intended for elementary school students. The development model used in this study refers to the Borg & Gall model [48] which has been modified into six main stages to suit the needs and scope of the research, namely:

1. preliminary study and needs analysis,
2. construction and indicator development,
3. item development,
4. content validation by experts,
5. limited testing, and
6. analysis of the validity and reliability of the instrument.

This model was chosen because it is flexible and systematic and has been widely used in developing educational instruments, especially those involving an integrated quantitative and qualitative approach.

2.2. Research Location and Participants

This study was conducted in six public elementary schools in West Sumatra Province, Indonesia, with diverse geographical and socio-economic characteristics. Schools were selected purposively, considering the willingness of the schools, the availability of teachers who applied the STEAM approach, and the representation of urban and semi-urban contexts.

The test subjects were 120 fifth-grade students from the six schools, selected using cluster random sampling based on class availability and parental consent. Fifth grade was chosen because students at this level generally have adequate basic literacy skills to understand the instrument items and are in the cognitive development stage towards the formal operational stage according to Piaget's theory [49, 50].

2.3. Research Instrument

The instrument developed was a non-test questionnaire with a 5-point Likert rating scale, ranging from "very inappropriate" to "very appropriate." This instrument aimed to measure five main aspects of 21st-century skills, namely:

1. Critical thinking,
2. Creativity,
3. Collaboration,
4. Communication, and
5. Digital literacy.

These five aspects are arranged based on the framework of the Partnership for 21st Century Learning [6] which is then contextualized with the STEAM (Science, Technology, Engineering, Arts, and Mathematics) approach in the indicator development process. Each aspect has several indicators reflecting student competencies in project-based learning or interdisciplinary problem-solving activities.

An example of a critical thinking item is: *"I can explain the reasons for choosing a solution in a group project."* Meanwhile, for the collaboration aspect, *"I respect my friends' opinions when developing an experiment plan."*

The initial instrument consisted of 38 items, which were then refined based on input from experts and the results of field trials.

2.4. Instrument Validation

Content validation was conducted by three experts with backgrounds in primary education, educational evaluation, and STEAM curriculum. The validation process was carried out using two approaches:

1. Content Validity Ratio (CVR) – used to assess the relevance of each item to the theoretical construct. Experts were asked to rate each item using a 3-point scale: "necessary," "necessary with revision," and "unnecessary." The CVR value was calculated using Lawshe's formula [51].
2. Content Validity Index (CVI) – to see the overall agreement on the quality of the instrument, particularly in terms of language clarity, readability, and cultural context appropriateness [52].

Items with a CVR value < 0.33 (for three experts) were eliminated or revised, while the overall CVI must be ≥ 0.80 for the instrument to be considered suitable for field testing.

2.5. Data Collection Procedure

Data were collected by distributing printed questionnaires to students who had obtained approval from the school and parents. The filling process was carried out in class with the assistance of teachers and researchers. The trial was conducted over two weeks in May 2025. All procedures were carried out per research ethics principles, including informed consent, data confidentiality, and voluntary participation.

2.6. Data Analysis Techniques

The data obtained were analyzed quantitatively to assess construct validity and instrument reliability. The analysis techniques used included:

1. Construct validity testing was conducted through Confirmatory Factor Analysis (CFA) using AMOS v24 software. CFA was used to verify the instrument's factor structure, which was developed based on the theoretical model of the five dimensions of 21st-century skills.
    The model goodness-of-fit criteria refer to the following values:
- Chi-square/df ≤ 3,
- CFI ≥ 0.90,
- TLI ≥ 0.90,
- RMSEA ≤ 0.08 [6].
2. Internal reliability testing was conducted using Cronbach's Alpha for each subscale, with a minimum reliability threshold of $\alpha \geq 0.70$ to indicate good internal consistency [7]. In addition, composite reliability (CR) and average variance extracted (AVE) were used to examine construct consistency and convergence. The entire analysis was done using IBM SPSS Statistics v26 software for initial descriptive analysis and AMOS v24 for CFA.

2.7. Access to Codes and Data

The final instruments and statistical analysis documentation are stored in digital form. If needed for replication or further study, the raw data and CFA analysis scripts can be accessed through an open repository linked in this article's appendix or additional references section.

3. Results

3.1. Content Validation Results by Experts

Content validation of the instrument was conducted by three experts with expertise in basic education, learning evaluation, and STEAM-based curriculum development. The experts were asked to evaluate 38 items of the initial instrument based on three main criteria: relevance to 21st-century skills constructs, readability by elementary school students, and contextual appropriateness with the STEAM approach. The validation process used a three-point scale: 1 = not necessary, 2 = necessary with revision, and 3 = necessary. This assessment was used to calculate each item's Content Validity Ratio (CVR), using the CVR formula developed by Lawshe [51]. In the context of three experts, the minimum acceptable CVR threshold was ≥ 0.33.

The validation results showed that of the 38 items developed, 32 items had a CVR value of ≥ 0.33 and were declared valid, while the other six items did not meet the requirements and were declared invalid in terms of content. The revision process was carried out on items declared invalid, either through editorial refinement or deletion of items deemed irrelevant. The total Content Validity Index (CVI) score for the entire instrument reached 0.89, reflecting a high level of agreement among experts on the overall quality of the instrument [52].

Details of the content validation results based on 21st-century skill dimensions can be seen in the following table:

Table 1.
Validation Results.

Skill Dimension	Initial Item Count	Valid Items (CVR ≥ 0.33)	Invalid Items	Average CVR	CVI per Dimension
Critical Thinking	7	6	1	0.78	0.90
Creativity	7	6	1	0.67	0.88
Collaboration	8	7	1	0.83	0.91
Communication	8	7	1	0.83	0.89
Digital Literacy	8	6	2	0.61	0.86
Total / Average	38	32	6	0.74	0.89

The data in Table 1 shows that the collaboration and communication dimensions have the highest average CVR value, namely 0.83, which indicates a very good level of content validity. Meanwhile, the digital literacy dimension has the lowest average CVR (0.61) and the most invalid items, namely, two items. This indicates that in the context of elementary school students, digital literacy may require a more specific language and context approach to be understood correctly.

As an illustration, the following are the results of content validation in the critical thinking dimension, which consists of seven initial items. Three experts and the CVR value evaluated each item and calculated the value. Six of the seven items were declared valid, while one had a CVR value below 0.33.

Table 2.
Example of Critical Thinking Dimension Validation.

Item Code	Brief Statement	Expert 1	Expert 2	Expert 3	CVR
BK-1	I can explain the reasons for choosing the solution	3	3	3	1.00
BK-2	I consider all ideas before making a decision	3	3	2	0.67
BK-3	I follow my friends' opinions without thinking twice	1	2	1	-0.33
BK-4	I can distinguish between facts and opinions in discussions	3	3	3	1.00
BK-5	I use logical reasoning when solving problems	3	2	3	0.67
BK-6	I ask questions when I don't understand the problem	3	3	3	1.00
BK-7	I draw conclusions based on the available information	3	2	2	0.33

From Table 2, item BK-3 was declared invalid because it had a CVR value of -0.33, and was removed from the instrument. Meanwhile, the other items were retained for further testing. A similar pattern also occurred in other dimensions, where invalid items generally had ambiguous wording or were not appropriate for the cognitive development context of elementary school students.

After the validation and revision process, a final instrument of 32 items was obtained, which were distributed proportionally across the five dimensions of 21st-century skills: six items for critical thinking, six items for creativity, seven items for collaboration, seven items for communication, and six items for digital literacy. These results indicate that the instrument has a balanced content structure and meets the criteria for strong content validity for use in measuring the 21st-century skills of elementary school students based on the STEAM approach.

3.2. Description of Field Test Results Data

The content-validated instrument, consisting of 32 items, was then tested on 120 fifth-grade students from six elementary schools in West Sumatra Province, Indonesia. This trial aimed to obtain an initial overview of data distribution and student response trends and detect possible item biases or item incompatibility with respondent characteristics. The completion process was carried out in a structured manner in the classroom, with assistance from teachers and researchers, and took between 30 and 40 minutes.

Preliminary data analysis was performed using descriptive statistics, including the mean and standard deviation, for each dimension of 21st-century skills. The results show that the average score per dimension ranges from 3.62 to 4.31 on a 5-point Likert scale, while the standard deviation values range from 0.51 to 0.68. These values indicate that, in general, student responses are in the positive category, and there are no indications of extreme deviations (outliers) or abnormal response patterns.

More specifically, the collaboration dimension recorded the highest average value (mean = 4.31) with a standard deviation of 0.51, indicating that students' collaboration skills have developed well. This is likely because group-based and project-based learning have become part of routine activities in elementary schools. Meanwhile, the digital literacy dimension showed the lowest mean score (mean = 3.62) and a relatively higher standard deviation (0.68), indicating variation in students' abilities in using digital technology. This is possible due to differences in access to digital devices in school and home environments and the suboptimal integration of technology in daily learning.

A summary of the descriptive data for the five dimensions of 21st-century skills is presented in the following table:

Table 3.
Results of the Field Test Calculations

21st Century Skill Dimensions	Number of Items	Mean	Standard Deviation (SD)	Response Category
Critical Thinking	6	3.98	0.61	High
Creativity	6	3.86	0.66	High
Collaboration	7	4.31	0.51	Very High
Communication	7	4.10	0.55	High
Digital Literacy	6	3.62	0.68	Fairly High

The data in Table 3 shows that almost all dimensions are in the "high" category, except for digital literacy, which is in the "fairly high" category. This shows that, in general, the elementary school students who were respondents have shown a positive trend towards the 21st-century skills measured, with a relatively better level of mastery in social aspects (collaboration and communication) than technological ones.

Furthermore, the standard deviation values, which tend to be low in all dimensions, indicate that the distribution of student responses is relatively even, without any high deviation from the mean. This reinforces the assumption that the items in the instrument are well-suited to the abilities and experiences of elementary school students. No extreme response patterns were found (e.g., all scores were "strongly agree" or "strongly disagree"), indicating that this instrument is quite sensitive in capturing variations in abilities between individuals.

These descriptive findings provide initial indications that the developed instrument has good content validity and can also be understood and used effectively by elementary school students, in accordance with their cognitive development and learning experience levels. This is an important foundation before further constructing a validity analysis through confirmatory factor analysis (CFA) and reliability testing.

3.3. Results of Construct Validity Analysis (CFA)

After the instrument was declared valid in content and field testing was conducted, the next step was to test construct validity using the Confirmatory Factor Analysis (CFA) approach within the Structural Equation Modeling (SEM) framework. This analysis was conducted using AMOS software version 24 to ensure that the instrument's factor structure was in accordance with the theoretical model that had been designed, namely 21st-century skills consisting of five main constructs: critical thinking, creativity, collaboration, communication, and digital literacy.

Several indicators developed into items, each consisting of 6–7 items, measure each construct. The CFA model tests the extent to which empirical data support the theoretical structure of the five dimensions, taking into account the goodness-of-fit indicators and factor loadings of each item.

The analysis results show that the measurement model has an excellent level of fit, with the following indicators:

- Chi-square/df = 2.36 (a value below 3 indicates model fit),
- Comparative Fit Index (CFI) = 0.94,
- Tucker Lewis Index (TLI) = 0.92,
- Root Mean Square Error of Approximation (RMSEA) = 0.063.

These values overall meet the model fit criteria suggested by Lawshe. Thus, the five-dimensional construct model developed is empirically consistent with the data obtained from the instrument trial.

In addition to model fit indicators, CFA also produced important information in the form of factor loadings for each item on its construct. High loading values indicate that the item represents the construct well. In this study, all factor loadings were in the range of 0.61 to 0.83, and all were significant at the $p < 0.01$ level. This shows that each item of the instrument substantially contributes to the measured 21st-century skills construct.

The following is a summary of the CFA factor loading results for each dimension:

Table 4.
Factor Loading Results.

21st Century Skills Dimension	Number of Items	Factor Loading Range	Average Factor Loadings	Significance
Critical Thinking	6	0.63 – 0.80	0.72	p < 0.01
Creativity	6	0.61 – 0.78	0.70	p < 0.01
Collaboration	7	0.68 – 0.83	0.76	p < 0.01
Communication	7	0.66 – 0.81	0.74	p < 0.01
Digital Literacy	6	0.62 – 0.79	0.71	p < 0.01

Based on Table 4, All constructs showed average factor loadings above 0.70, indicating a strong relationship between the items and the constructs. Loadings greater than 0.60 are considered adequate in SEM analysis, and greater than 0.70 indicates that each item can represent its construct consistently.

Furthermore, no items with loadings below 0.60 were found, which are usually candidates for elimination. Therefore, all 32 items were deemed feasible to retain in the final model and can be used to measure the 21st-century skills of elementary school students in the context of STEAM-based learning.

These findings reinforce that the five-dimensional instrument structure developed is not only content-valid but also has excellent construct validity based on empirical testing. This construct validity provides a strong foundation for trusting the instrument's measurement results and using them to assess student competencies more holistically and evidence-based.

3.4. Reliability Test Results

Reliability analysis aims to measure the level of internal consistency of the developed instrument to ensure that each item in a construct dimension provides stable and consistently correlated results. In this study, reliability testing was conducted on the five main constructs of 21st-century skills using two complementary statistical approaches, namely:

- Cronbach's Alpha – measures internal consistency between items in a single construct.
- Composite Reliability (CR) and Average Variance Extracted (AVE) measure the overall reliability of the construct in the CFA (Confirmatory Factor Analysis) model.

A Cronbach's Alpha value above 0.70 is considered adequate, indicating that each dimension has strong internal consistency. Meanwhile, a CR value is considered good if it is ≥ 0.70, and an AVE value of ≥ 0.50 indicates that the construct has fairly good convergent validity.

The reliability test results for the five dimensions are presented in the following table:

Table 5.
Reliability Test Results for the 21st Century Skills Instrument.

21st Century Skills Dimension	Cronbach's Alpha	Composite Reliability (CR)	Average Variance Extracted (AVE)
Critical Thinking	0.81	0.85	0.57
Creativity	0.78	0.82	0.54
Collaboration	0.83	0.87	0.61
Communication	0.80	0.84	0.56
Digital Literacy	0.76	0.81	0.52

Based on Table 5 it can be seen that all dimensions have a Cronbach's Alpha value above 0.76, which means that each item in each construct has a consistent and reliable correlation for use in repeated measurements. The collaboration dimension shows the highest reliability (α = 0.83; CR = 0.87), confirming that students tend to respond consistently to items in this construct.

The Composite Reliability (CR) values also show strong results, ranging from 0.81 to 0.87, reinforcing the confidence that each construct has been accurately measured in the structural model. Furthermore, the AVE values for all dimensions were above 0.50, indicating that most of the variance in each construct was successfully explained by its constituent items, thus meeting the criteria for convergent validity.

These reliability test results reinforce previous findings that the developed instrument has high measurement quality in terms of construct validity and internal consistency. Thus, this instrument is suitable for use to measure 21st-century skills, particularly in the context of STEAM learning at the elementary school level. This instrument is not only statistically accurate but also stable in measuring the important aspects students need to face the challenges of the 21st century.

Based on all stages of analysis, the 21st-century skills measurement instrument developed in this study has met the criteria for theoretical and empirical feasibility. The final instrument consists of 32 statements distributed proportionally across the five dimensions of 21st-century skills: critical thinking, creativity, collaboration, communication, and digital literacy. The results of expert content validation showed a very high level of agreement with a Content Validity Index (CVI) value of 0.89. At the same time, item selection and revision were based on the Content Validity Ratio (CVR) value. Construct validity testing through Confirmatory Factor Analysis (CFA) confirmed that the five-dimensional model has good statistical validity (CFI = 0.94; RMSEA = 0.063), supported by significant and strong factor loadings. In addition, internal reliability testing showed that all dimensions had a high level of consistency with Cronbach's Alpha values between 0.76 and 0.83, and Composite Reliability (CR) between 0.81 and 0.87. Thus, this instrument can be declared valid and reliable for measuring the 21st-century skills of elementary school students, particularly in the context of STEAM-based learning.

4. Discussion

This study produced a measurement instrument for 21st-century skills specifically designed for elementary school students in the context of STEAM-based learning. The instrument's content validity, construct validity, and reliability have been tested and meet strong scientific criteria. Overall, these findings support the working hypothesis that 21st-century skills in elementary school students can be measured validly and reliably if developed with the appropriate pedagogical and theoretical approaches.

High content validity, as indicated by a Content Validity Index (CVI) value of 0.89, indicates a very good level of agreement among experts regarding the relevance and readability of the items [53]. This aligns with Lawshe [51] statement, which emphasizes the importance of expert involvement in assessing the suitability of items to the construct. The CVR value used to select items also shows that each instrument item was developed based on the principle of concept representativeness [54].

The construct validity test through Confirmatory Factor Analysis (CFA) shows that the five-dimensional model of 21st-century skills—critical thinking, creativity, collaboration, communication, and digital literacy—has a good level of model fit. The CFI value of 0.94 and RMSEA of 0.063 are within the range suggested by Hu and Bentler [55] which states that CFI ≥ 0.90 and RMSEA ≤ 0.08 indicate a statistically suitable model. By Martynova, et al. [56] emphasizes that a well-fitting measurement model is an important basis for construct validity, demonstrating the relationship between theory and empirical data.

The dimension with the strongest consistency is collaboration, with the highest factor loading and internal reliability of 0.83. This indicates that aspects of cooperation and social interaction are highly developed in elementary school students, which aligns with Vygotsky's view of the importance of social interaction in children's cognitive development Nguyen, et al. [57]. Zainil, et al. [19] state that the STEAM approach naturally enhances collaboration skills through project completion that requires student coordination [19].

In contrast, the digital literacy dimension showed the lowest average score (mean = 3.62), indicating that elementary school students still have low mastery of technological skills. This reinforces the findings of Hsu, et al. [58] who revealed that digital literacy at the elementary level is highly dependent on access to and integration of technology in learning Hsu, et al. [58]. Peng and Yu [59] emphasizes that digital literacy is a technical ability and includes critical thinking skills in using digital information [59].

In general, all dimensions showed Cronbach's Alpha reliability values above 0.76 and Composite Reliability (CR) above 0.81, exceeding the minimum limit of 0.70 recommended by Kennedy [60]. Haji-Othman, et al. [61] also emphasize

that a CR value above 0.80 indicates strong construct reliability, and an AVE value greater than 0.50 reflects good convergent validity [61].

The instruments developed in this study also made an important contribution to the pedagogical approach. As suggested by Ramadhani, et al. [29] the integration of the STEAM approach in the development of instrument indicators provided a more realistic and interdisciplinary learning context. The STEAM approach has been proven to develop critical thinking, creativity, and collaboration skills through project-based activities [62].

From a theoretical perspective, this study's results expand the model for measuring 21st-century skills, which was previously applied more to secondary or higher education levels [63]. These findings also support the argument that 21st-century skills can and should be developed from the elementary level, with a learning approach appropriate for students' developmental age [32].

The practical implications of this research are extensive. First, teachers can use this instrument as a diagnostic tool to assess students' strengths and weaknesses in the five dimensions of 21st-century skills. The assessment of these skills must be formative so that it can guide adaptive learning strategies [64]. Second, schools can use the measurement results to make decisions in curriculum planning and developing project-based learning (PBL) or STEM/STEAM programs.

Third, the government and education policymakers can use this instrument to formulate data-based policies to strengthen character education and future competencies. OECD [65] states that 21st-century education systems must be oriented toward developing transferable complex competencies, not merely content mastery [66].

However, as with other studies, this study has several limitations that must be considered. First, the pilot test was only conducted on fifth-grade students in six elementary schools in West Sumatra Province. Although it covers contextual variations, these results cannot be generalized nationally [67]. Second, the instrument was still in the form of a self-assessment questionnaire that was highly dependent on the respondents' understanding, which at the elementary school age could be biased or influenced by the desire to "please" the teacher or researcher. Third, the analysis is still limited to a covariance-based CFA approach. In the future, using the Rasch model or multi-group CFA approach can provide more detailed information about construct invariance between groups [68]. Based on these results and limitations, further research is recommended to test the instrument on a broader scale and across regions. In addition, a digital version of this instrument needs to be developed to be used efficiently in online or hybrid learning systems. Finally, integrating this instrument with performative assessments such as STEAM-based project rubrics can result in a more authentic and comprehensive assessment approach [69].

Overall, this instrument can make an important contribution to the development of educational assessments relevant to the needs of the 21st century. With its proven validity and reliability, as well as its STEAM-based contextual approach, this instrument has the potential to be used in various contexts of learning and national basic education evaluation.

5. Conclusion

Using a STEAM-based learning approach, this study produced a valid and reliable measurement instrument for 21st-century skills for elementary school students. The final instrument consists of 32 items reflecting five main dimensions—critical thinking, creativity, collaboration, communication, and digital literacy—with high content validity (CVI = 0.89), an appropriate CFA model (CFI = 0.94, RMSEA = 0.063), and strong internal reliability (α = 0.76–0.83; CR = 0.81–0.87). These results indicate that 21st-century skills can be measured validly and consistently from the elementary school level if developed using a contextual and pedagogically integrated approach. This instrument serves as an evaluation tool and as a foundation for developing learning that is more relevant to future demands. However, limitations such as geographical constraints and self-report-based measurement methods must be considered. Therefore, further testing on a larger scale, the development of a digital version, and integration with performative assessments are recommended to strengthen the quality and utilization of this instrument in the national education system.

References

[1]   N. Akcanca, "21st century skills: the Predictive role of attitudes regarding stem education and problem-based learning," *International Journal of Progressive Education,* vol. 16, no. 5, pp. 443-458, 2020.

[2]   C. Martinez, "Developing 21st century teaching skills: A case study of teaching and learning through project-based curriculum," *Cogent Education,* vol. 9, no. 1, p. 2024936, 2022. https://doi.org/10.1080/2331186X.2021.2024936

[3]   I. Tuomi, "Artificial intelligence, 21st century competences, and socio-emotional learning in education: More than high-risk?," *European Journal of Education,* vol. 57, no. 4, pp. 601-619, 2022. https://doi.org/10.1111/ejed.12531

[4]   I. Rahmania, "Project based learning (PjBL) learning model with STEM approach in natural science learning for the 21st century," *Budapest International Research and Critics Institute (BIRCI-Journal): Humanities and Social Sciences,* vol. 4, no. 1, pp. 1161-1167, 2021. https://doi.org/10.33258/birci.v4i1.1727

[5]   Y. Erita, Y. Miaz, J. Jupriani, S. Hevria, and R. Fauzi, "Using virtual reality to enhance twenty-first-century skills in elementary school students: A systematic literature review," *Open Education Studies,* vol. 6, no. 1, p. 20240030, 2024. https://doi.org/10.1515/edu-2024-0030

[6]   T. J. Kennedy and C. W. Sundberg, *21st century skills*. Cham: Springer, 2020, pp. 479-496.

[7]   M. Jannah, L. D. Prasojo, and M. A. Jerusalem, "Elementary school teachers' perceptions of digital technology based learning in the 21st century: Promoting digital technology as the proponent learning tools," *Al Ibtida: Jurnal Pendidikan Guru MI,* vol. 7, no. 1, pp. 1-18, 2020.

[8]   T. Langenfeld, J. Burstein, and A. A. von Davier, "Digital-first learning and assessment systems for the 21st century," in *Frontiers in Education*, 2022, vol. 7, p. 857604.

[9]   A. Mahpudz and A. Palimbong, "Tolerance learning to develop students social skills in the 21st Century," presented at the International Conference on Social Studies, Globalisation and Technology (ICSSGT 2019), 2020.

[10] R. Tytler, "STEM education for the twenty-first century," *Integrated Approaches to STEM Education: An International Perspective,* pp. 21-43, 2020. https://doi.org/10.1007/978-3-030-52229-2_3

[11] S. K. W. Chu, R. B. Reynolds, N. J. Tavares, M. Notari, and C. W. Y. Lee, *21st century skills development through inquiry-based learning: From theory to practice*. Singapore: Springer Singapore, 2017.

[12] Y. Yerizon, A. Arnellis, S. Suherman, I. M. Arnawa, C. Sa'dijah, and L. Anwar, "Developing adaptive digital book (ADB) in enhancing students' numeracy literacy ability," *Infinity Journal,* vol. 14, no. 3, pp. 571-586, 2025. https://doi.org/10.22460/infinity.v14i3.p571-586

[13] D. R. Asnawi, T. M. Sahudra, and A. Kiswanto, "Primary school teachers' perceptions of differentiation learning," in *Proceedings of the 6th International Conference on Current Issues in Education (ICCIE) 2023*, 2024, vol. 847, p. 249.

[14] E. Noviana, O. Kurniaman, S. Munjiatun, and S. D. Nirmala, "Why do primary school students need disaster mitigation knowledge?(Study of the use of Koase Comics in primary schools)," *International Journal of Scientific & Technology Research,* vol. 8, no. 11, pp. 216-221, 2019.

[15] K. Tsarava *et al.*, "A cognitive definition of computational thinking in primary education," *Computers & Education,* vol. 179, p. 104425, 2022. https://doi.org/10.1016/j.compedu.2021.104425

[16] E. Yayuk and A. R. As' ari, "Primary school students' creative thinking skills in mathematics problem solving," *European Journal of Educational Research,* vol. 9, no. 3, pp. 1281-1295, 2020.

[17] J. Ahmad and N. M. Siew, "Curiosity towards STEM Education: A Questionnaire for Primary School Students," *Journal of Baltic Science Education,* vol. 20, no. 2, pp. 289-304, 2021. https://doi.org/10.33225/jbse/21.20.289

[18] D. Triana, Y. U. Anggraito, and S. Ridlo, "Effectiveness of environmental change learning tools based on STEM-PjBL towards 4C skills of students," *Journal of Innovative Science Education,* vol. 9, no. 2, pp. 181-187, 2020.

[19] M. Zainil, A. K. Kenedi, Rahmatina, T. Indrawati, and C. Handrianto, "The influence of STEM-based digital learning on 6C skills of elementary school students," *Open Education Studies,* vol. 6, no. 1, p. 20240039, 2024.

[20] D. Digna and C. Widyasari, "Teachers' perceptions of differentiated learning in merdeka curriculum in elementary schools," *International Journal of Elementary Education,* vol. 7, no. 2, pp. 255-262, 2023. https://doi.org/10.23887/ijee.v7i2.54770

[21] I. Sujarwati, "The differentiated learning strategy in implementation merdeka belajar curriculum to improve students' learning outcomes of English lesson in elementary school," *Cendikia: Media Jurnal Ilmiah Pendidikan,* vol. 13, no. 5, pp. 724-733, 2023.

[22] N. Pourjabali and N. A. Almasi, "A comparative study of post-marriage nationality of women in legal systems of different countries," *International Journal of Multicultural and Multireligious Understanding,* no. 2, p. 3, 2015.

[23] A. Fauzan, Y. Harisman, and F. Sya'bani, "Local instructional theory of probability topics based on realistic mathematics education for eighth-grade students," *Journal on Mathematics Education,* vol. 13, no. 4, pp. 703-722, 2022. https://doi.org/10.22342/jme.v13i4.pp703-722

[24] A. Fauzan, Y. Harisman, G. Boriboon, H. Syaputra, and H. Hafizatunnisa, "Enhancing mathematical literacy: A comparative study of mathematics learning curricula in Thailand and Indonesia," *Infinity Journal,* vol. 14, no. 3, pp. 753-780, 2025. https://doi.org/10.22460/infinity.v14i3.p753-780

[25] E. Musdi, A. R. As'ari, Y. Harisman, H. Syaputra, and K. A. Hevardani, "Student's creative thinking based on study level, learning style, gender, and combination of the three," *International Journal of Evaluation and Research in Education,* vol. 13, no. 3, pp. 1591-1601, 2024. https://doi.org/10.11591/ijere.v13i3.27936

[26] A. R. Saavedra and V. D. Opfer, "Learning 21st-century skills requires 21st-century teaching," *Phi Delta Kappan,* vol. 94, no. 2, pp. 8-13, 2012. https://doi.org/10.1177/003172171209400203

[27] H. Kim, "Education system alignment for 21st century skills: Focus on assessment," 2019. https://www.researchgate.net/publication/330740772

[28] H. Syarifuddin and B. Atweh, "The use of activity, classroom discussion, and exercise (ACE) teaching cycle for improving students' engagement in learning elementary linear Algebra," *European Journal of Science and Mathematics Education,* vol. 10, no. 1, pp. 104-138, 2022.

[29] D. Ramadhani, A. K. Kenedi, M. F. Rafli, H. Harahap, L. C. Aritonang, and A. Andini, "Analysis of the need for the development of STEM-PjBL learning models for elementary school students," in *Proceedings of the 1st International Conference on Early Childhood Education in Multiperspective (ICECEM 2022, 26 November 2022, Purwokerto, Central Java, Indonesia). Purwokerto, Central Java, Indonesia: EAI,* 2023.

[30] D. Ramadhani, A. K. Kenedi, M. F. Rafli, and C. Handrianto, "Advancement of STEM-based digital module to enhance HOTS of prospective elementary school teachers," *Jurnal Pendidikan Progresif,* vol. 12, no. 2, pp. 981-993, 2022. https://doi.org/10.23960/jpp.v12.i2.202245

[31] A. Arwin, A. Kenedi, Y. Anita, C. Hamimah, and M. Zainil, "STEM-based digital disaster learning model for disaster adaptation ability of elementary school students," *International Journal of Evaluation and Research in Education ISSN,* vol. 13, no. 5, pp. 3248-3259, 2024. https://doi.org/10.11591/ijere.v13i5.29616

[32] M. Zainil and A. K. Kenedi, "Advancement of STEM-based e-student worksheet to enhance the HOTS of elementary school students," *Journal of Education Technology,* vol. 6, no. 3, pp. 478-488, 2022.

[33] M. Zainil, A. K. Kenedi, T. Indrawati, and C. Handrianto, "The influence of a stem-based digital classroom learning model and high-order thinking skills on the 21st-century skills of elementary school students in Indonesia," *Journal of Education and E-Learning Research,* vol. 10, no. 1, pp. 29-35, 2023. https://doi.org/10.20448/jeelr.v10i1.4336

[34] J. D. Basham, M. Israel, and K. Maynard, "An ecological model of STEM education: Operationalizing STEM for all," *Journal of Special Education Technology,* vol. 25, no. 3, pp. 9-19, 2010. https://doi.org/10.1177/016264341002500303

[35] C. L. Bowen, S. Heubach, and J. J. Dong, "Eco-STEM tools: Changing the culture of teaching and learning in STEM," presented at the 2022 IEEE Frontiers in Education Conference (FIE), 2022.

[36] F. Kartini, A. Widodo, and N. Winarno, "STEM project-based learning on student's STEM literacy: the case of teaching earth layer and disaster," *Journal of Physics: Conference Series,* vol. 1806, no. 1, p. 012221, 2021. https://doi.org/10.1088/1742-6596/1806/1/012221

[37] A. Agusty, F. Alifteria, and M. Anggaryani, "STEM in disaster learning media: A literature review," *Journal of Physics: Conference Series,* vol. 2110, no. 1, p. 012016, 2021. https://doi.org/10.1088/1742-6596/2110/1/012016

[38] S. Hanif, A. F. C. Wijaya, and N. Winarno, "Enhancing students' creativity through stem project-based learning," *Journal of science Learning,* vol. 2, no. 2, pp. 50-57, 2019. https://doi.org/10.17509/jsl.v2i2.13271

[39] D. Henriksen, "Full STEAM ahead: Creativity in excellent STEM teaching practices," *The Transdisciplinary STEAM+ Journal,* vol. 1, no. 2, p. 15, 2013. https://doi.org/10.5642/steam.20140102.15

[40] S. A. Buriro, M. A. Mirjat, R. L. Pathan, I. Chandio, A. A. Lashari, and H. Gul, "Eco-Friendly pedagogies for STEM education: A review," *Journal of Namibian Studies: History Politics Culture,* vol. 34, pp. 3018-3044, 2023.

[41] K. Septaria, B. A. Dewanti, and M. I. El Afidah, "Development of module disaster mitigation based on STEM for secondary schools," *Jurnal Penelitian Pendidikan IPA,* vol. 5, no. 2, pp. 61-68, 2020.

[42] Y. Fitria and A. K. Kenedi, "The analysis of online learning process in science for elementary school during Covid-19 pandemic," *AIP Conference Proceedings,* vol. 2468, no. 1, p. 060025, 2022.

[43] Y. Helsa and A. K. Kenedi, "Edmodo-based blended learning media in learning mathematics," *Journal Of Teaching and Learning in Elementary Education,* vol. 2, no. 2, pp. 107-117, 2019. https://doi.org/10.33578/jtlee.v2i2.7416

[44] T. Sahudra, A. K. Kenedi, and A. Asnawi, "Preliminary study of ethno-social learning advancement to enhance the value of Pancasila student profiles," in *Proceedings of the 2nd International Conference on Education, Humanities, Health and Agriculture (ICEHHA 2022, 21–22 October 2022, Ruteng, Flores, Indonesia) (pp. ...). Ruteng, Flores, Indonesia: EAI,* 2023.

[45] N. Astimar, Y. Helsa, and A. Kenedi, "The review of utilization in natural environment by the teacher as a source for learning science at elementary school in low class," *Journal of Physics: Conference Series,* vol. 1424, no. 1, p. 012036, 2019. https://doi.org/10.1088/1742-6596/1424/1/012036

[46] A. Asnawi, A. Kenedi, and R. Fransyaigu, "The influence of digital classroom model using local wisdom towards elementary school student's learning motivation and learning independence in Indonesia," *European Alliance for Innovation,* 2022. https://doi.org/10.4108/eai.26-11-2022.2339527

[47] Y. Anita, A. Wa, A. U. Akmal, and A. K. Kenedi, "Science electronic teaching materials based on social-emotional learning as an attempt to enhance the value of pancasila student profiles," presented at the International Conference on Educational Technology, 2022.

[48] K. A. Aka, "Integration Borg & Gall (1983) and Lee & Owen (2004) models as an alternative model of design-based research of interactive multimedia in elementary school," *Journal of Physics: Conference Series,* vol. 1318, no. 1, p. 012022, 2019.

[49] F. Niklas and W. Schneider, "Home learning environment and development of child competencies from kindergarten until the end of elementary school," *Contemporary Educational Psychology,* vol. 49, pp. 263-274, 2017. https://doi.org/10.1016/j.cedpsych.2017.03.006

[50] E. G. Merritt, S. B. Wanless, S. E. Rimm-Kaufman, C. Cameron, and J. L. Peugh, "The contribution of teachers' emotional support to children's social behaviors and self-regulatory skills in first grade," *School Psychology Review,* vol. 41, no. 2, pp. 141-159, 2012. https://doi.org/10.1080/02796015.2012.12087517

[51] C. H. Lawshe, "A quantitative approach to content validity," *Personnel Psychology,* vol. 28, no. 4, pp. 563-575, 1975.

[52] F. Madadizadeh and S. Bahariniya, "Tutorial on how to calculating content validity of scales in medical research," *Perioperative Care and Operating Room Management,* vol. 31, p. 100315, 2023. https://doi.org/10.1016/j.pcorm.2023.100315

[53] D. F. Polit and C. T. Beck, *Nursing research: Principles and methods*. Philadelphia, PA: Lippincott Williams & Wilkins, 2004.

[54] R. F. DeVellis and C. T. Thorpe, *Scale development: Theory and applications*. Thousand Oaks, CA: SAGE Publications, 2021.

[55] L. t. Hu and P. M. Bentler, "Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives," *Structural Equation Modeling: A Multidisciplinary Journal,* vol. 6, no. 1, pp. 1-55, 1999. https://doi.org/10.1080/10705519909540118

[56] E. Martynova, S. G. West, and Y. Liu, "Review of principles and practice of structural equation modeling," *Structural Equation Modeling: A Multidisciplinary Journal,* vol. 25, no. 2, pp. 325-329, 2018. https://doi.org/10.1080/10705511.2017.1401932

[57] M.-H. Nguyen, V.-P. La, T.-T. Le, and Q.-H. Vuong, "Introduction to bayesian mindsponge framework analytics: An innovative method for social and psychological research," *MethodsX,* vol. 9, p. 101808, 2022. https://doi.org/10.1016/j.mex.2022.101808

[58] T.-C. Hsu, Y.-S. Chang, M.-S. Chen, I.-F. Tsai, and C.-Y. Yu, "A validity and reliability study of the formative model for the indicators of STEAM education creations," *Education and Information Technologies,* vol. 28, no. 7, pp. 8855-8878, 2023.

[59] D. Peng and Z. Yu, "A literature review of digital literacy over two decades," *Education Research International,* vol. 2022, no. 1, p. 2533413, 2022. https://doi.org/10.1155/2022/2533413

[60] I. Kennedy, "Sample size determination in test-retest and Cronbach alpha reliability estimates," *British Journal of Contemporary Education,* vol. 2, no. 1, pp. 17-29, 2022.

[61] Y. Haji-Othman, M. S. Sheh Yusuff, and M. N. Hussain, "Data analysis using Partial Least Squares Structural Equation Modeling (PLS-SEM) in conducting quantitative research," *International Journal of Academic Research in Business and Social Sciences,* vol. 14, no. 10, pp. 2380–2388, 2024.

[62] N. Pratiwi, F. A. Putri, and M. Ali, "The effect of stem approach in problem-based learning for increasing students' problem-solving ability in the topic of environmental pollution," *KnE Social Sciences,* 2024. https://doi.org/10.18502/kss.v9i13.16032

[63] A. Koculu, M. S. Topcu, and A. Ciftci, "The effect of STEM education on pre-service science teachers' perceptions of 21st century skills and competences and problem solving skills," *Open Journal for Educational Research,* vol. 6, no. 2, pp. 165-172, 2022.

[64] M. Suhaini, A. Ahmad, and N. M. Bohari, "Assessments on Vocational Knowledge and Skills: A Content Validity Analysis," *European Journal of Educational Research,* vol. 10, no. 3, pp. 1529-1540, 2021.

[65] OECD, *Policy coherence for sustainable development 2018: Towards sustainable and resilient societies*. Paris, France: OECD Publishing, 2018.

[66] M. Llamas-Nistal, M. J. Fernández-Iglesias, J. González-Tato, and F. A. Mikic-Fonte, "Blended e-assessment: Migrating classical exams to the digital world," *Computers & Education,* vol. 62, pp. 72-87, 2013.

[67] E. Colomo-Magaña, R. Soto-Varela, J. Ruiz-Palmero, and M. Gómez-García, "University students' perception of the usefulness of the flipped classroom methodology," *Education Sciences,* vol. 10, no. 10, p. 275, 2020.

[68]   H. C. Huang, Y. M. Tseng, Y. C. Chen, P. Y. Chen, and H. Y. Chiu, "Diagnostic accuracy of the Clinical Dementia Rating Scale for detecting mild cognitive impairment and dementia: A bivariate meta-analysis," *International Journal of Geriatric Psychiatry,* vol. 36, no. 2, pp. 239-251, 2021.

[69]   A. Asrial *et al.*, "E-Assessment for character evaluation in elementary schools," *Qubahan Academic Journal,* vol. 4, no. 3, pp. 806-822, 2024.